A Note on Performance Metrics for Warehouses

Kevin R. Gue

Department of Systems Management

Naval Postgraduate School

Monterey, CA 93943

krgue@nps.navy.mil

December 2, 1999

Abstract

We discuss two common performance metrics for a warehouse, and show how they encourage perverse or inefficient operational behavior. We propose a third metric and argue that it promotes rational work behavior and improved customer service.

1 Two performance metrics

A warehouse performs three major functions: receiving, storage, and shipping. Of the three, shipping is often considered most important because it is the function that a customer sees. Customers rarely care how efficiently a distribution center receives or stores its inventory; they want their goods as quickly as possible and in good condition.

To know how well a warehouse is performing the shipping function, managers construct *metrics*, which are most often numerical representations of data collected in the warehouse. Two common metrics are average days delayed and average cycle time. We evaluate each metric according to the following standards:

- Does it distinguish between warehouses well at the margin?
- Is it relatively independent of the order arrival stream?
- Does it encourage reasonable workforce planning? and
- Does better performance necessarily mean better customer service?

1.1 Average Days Delayed

Average days delayed is the average number of days required to ship a item, where each observation is rounded down. For example, an item that ships on the same day it arrived has zero days delay; an item that ships anytime during the next day has one day delay, and so on. For this measure, a *day* is defined by the clock; thus an item arriving at 0500 arrives the same day as one arriving at 2300, although it has a much greater chance of shipping the same day (and achieving zero days delay). Figure 1 illustrates three orders having the same 1-day delay time.

Note that to say that an item "has shipped" can different meanings, depending on the transportation mode. For example, if the warehouse uses common carriers such as UPS or FedEx, "shipped" may mean that it is ready for pickup, even though the carrier will not arrive for several hours. For a distribution center that controls its own fleet of trucks, "shipped" may mean that the truck has left the dock. For our purposes, we use the first definition; that is, we consider an item to have shipped when it is available for transporting.

The intent behind average days delayed is to push workers to ship items on the same day they are received, thus giving customers the advantage of receiving their orders sooner.

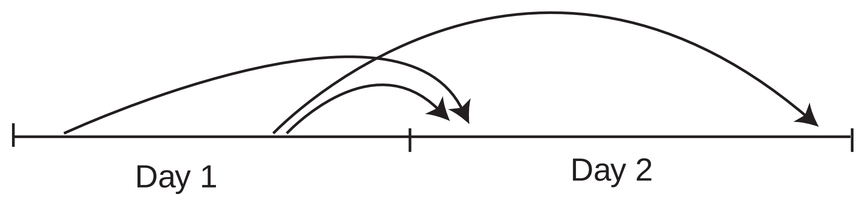

Figure 1: Three equivalent orders according to the average days delayed metric.

As a performance metric, average days delayed has several weaknesses. First, it does not record marginal performance for issues meeting the target number of days. For example, consider a warehouse that finishes half of its issues the same day they are received and the second half at 2000 the next day. A second warehouse finishes half the first day and half by 0800 the next day. The average days delayed metric views these warehouses as identical.

Second, the performance of the warehouse depends directly on the arrival times for orders, making it difficult to compare performance of different warehouses using the same metric. For example, a warehouse having a large influx of orders late in the day will necessarily perform poorly compared to one having a large influx earlier in the day. This is an important point, because one of the primary uses of metrics by managers is to compare different warehouses in the firm.

Third, the measure could cause supervisors to unnecessarily schedule irregular work hours, thus increasing labor costs. For example, if a large number of orders typically arrive late in the day, managers might schedule a late shift at higher labor cost per hour. But unless the late shift allows those orders to be transported on an earlier route, the customer sees no advantage.

Fourth, improving average days delayed does not necessarily improve customer service. For example, if there is a large surge late in the day, we might bring in a large afternoon shift to process all those orders before midnight, thus reducing average days delay; but if those orders still ship the next day, there is no advantage to the customer.

1.2 Average Cycle Time

The cycle time of an order is the difference between the time it arrives at the warehouse and the time it is ready for shipping. The average cycle time metric records the average value for all orders.

This metric does away with the first weakness above; that is, it better acknowledges marginal performance differences. A warehouse completing all orders in 8 hours every day has better performance than one completing all orders in 14 hours.

Average cycle time is also less affected by the timing of order arrivals. Two warehouses having a large influx of orders but at different times of the day should fare the same in average cycle time, although a later surge may require higher cost of labor.

Unfortunately, average order cycle time can cause managers to construct perverse workforce schedules. For example, if a surge of orders arrives overnight, there is an incentive to bring workers in during the mid-shift in order to get those orders processed, even though the orders cannot be shipped until the FedEx pickup late that afternoon. Those orders might have been processed just as well during normal work hours.

Finally, improving average order cycle time does not necessarily improve customer service. Reducing the average cycle time from 8 to 7 hours will have no effect unless some orders actually ship earlier. If the result is simply that orders spend more time on the shipping dock waiting for the carrier, then the customer sees no difference.

1.3 Why these measures fail to improve customer service

Average days delayed and average order cycle time fail to necessarily improve customer service because they are *internally-focused* metrics. They measure to different degrees how long it takes an order to flow through the warehouse, without regard to how soon the customer sees his order. In fact, both measures *tend to* improve customer service, but only as a side effect.

These metrics ignore the larger context of how a supply chain responds to customer orders. Warehouses and manufacturing sites can respond to demand continuously, where orders flow from the facility in a stream. Transportation is different because of the need for economies of scale—orders are served in batches, such as truckloads, whether they are constructed by the warehouse or by the carrier. Batching necessarily introduces a periodic nature to the distribution system, and it is this batching that the two measures ignore.

2 A customer-focused metric

To establish a truly customer-focused performance metric, we must account for the batching that occurs in transportation. We propose a *Percent making Cut-Off* (PCO) metric that records the fraction of orders arriving before an established cut-off time that make the next shipment cycle.

For example, suppose a carrier such as UPS or FedEx makes a pickup at 1700 every day, and we establish a cut-off time of 1500. The PCO metric would record the fraction of orders arriving before 1500 every day that ship at 1700 the same day. Those missing the mark would have arrived before 1500 the following day and counted in that day's metric as well. The idea is to ship as many orders as possible on the same day. This is the same goal aimed at by average days delayed, but here we have a sensibly defined "day".

Let's see how the PCO metric measures up according to our standards: The PCO metric does not distinguish between warehouses at the margin — the warehouse gains nothing by processing items early — but we contend that it does not matter because the warehouse finishing early does not improve customer service. (In the case of average days delayed, it could.) Performance on the PCO metric is definitely a function of the pattern of order arrivals: a warehouse having a surge close to the cut-off time will tend to perform poorly compared to one with an earlier surge. This is a weakness of the metric in that it makes cross-warehouse comparisons difficult.

The PCO metric drives rational workforce scheduling. It gives managers the incentive to plan in such a way that as much work as possible is out of the way prior to the cut-off time. Because the last minute rush happens at the same time (or times) every day, managers can free up resources to handle the increased load. Workers are working when the customer needs them to be.

Finally, improvements on the PCO metric result in better customer service by definition. Moreover, cut-off times provide the warehouse with an effective marketing tool. Because customer service is fundamentally about meeting or exceeding customer expectations, firms are at a disadvantage when the customer does not know what to expect or expects more than the firm can deliver. The PCO metric allows the warehouse to publish the cut-off time, clarifying expectations and improving customer satisfaction. For example, if a customer submits an order an hour after the cut-off time, he should have little or no expectation that his order will ship that day. If the warehouse happens to get the order out, the customer's expectation is exceeded; if not, it is met.

PC Connection, a mail-order computer distributor, publishes its next-day delivery cut-off on the front of its catalog. If they receive the order by 0200 Eastern time, the customer is to expect delivery the next (actually, the same) day.

Finally, we note that managers can modify the PCO metric to apply to multiple shipment types and transportation modes. For example, if shipping a certain type of bulk item takes extra time and resources, managers might establish the cut-off to be a day or two following the order.